

CO328 Deep Learning

MLR Meme Captioning Challenge, April 2024

https://drive.google.com/drive/folders/1zo5VhbfVk3dKQDm_mavg5a_eIJD9MKEf?usp=sharing

Overview

Memes have emerged as a pervasive medium for internet users to convey their ideas and sentiments through visual metaphors. Comprehending memes necessitates the ability to identify and interpret these visual metaphors in conjunction with the accompanying text, often requiring contextual knowledge and reasoning skills. We introduce the task of meme captioning, which aims to generate descriptive captions for memes by leveraging both the visual and textual components present in the meme.

The task of meme captioning presents a unique challenge in the field of multimodal machine learning, as it requires the integration of computer vision and natural language processing techniques. Specifically, the model must be capable of understanding the visual metaphor depicted in the meme image, while also considering the existing text within the meme and drawing upon relevant background knowledge to generate an appropriate and coherent caption.

This task has practical applications in areas such as content moderation, where automatically generated captions could aid in the detection and filtering of offensive or inappropriate memes. Additionally, it could contribute to the development of more advanced conversational agents that can effectively comprehend and respond to meme-based communication.

Title: one of them is my alt



Caption: Meme poster appreciates their only two followers and one of them is their alternative account.

Description

Your objective is to develop a multimodal model capable of generating meme captions. The model will take two inputs: an image (the meme) and a set of image captions (`img_captions`). Based on these inputs, the model must predict an appropriate `meme_caption` that encapsulates the visual metaphor and contextual information present in the meme image, while also incorporating relevant semantic information from the provided image captions.

The multimodal nature of this task necessitates the integration of computer vision and natural language processing techniques.

The model must be able to effectively understand and interpret the visual content of the meme image, as well as process and comprehend the textual information from the image captions. Subsequently, it must leverage this multimodal understanding to generate a coherent and contextually relevant meme caption.

Input

- * The Meme
- * Image Captions

Output

- * Meme Caption

About Dataset

This dataset has been created by scraping memes from reddit ('r/Memes') sub.

Files and Folders

- * `**memes-trainval.json**` - the training set
- * `**memes-->memes**` - Collection of all images (train, val, test)
- * `**memes-test.json**` - the Testing set (*****WILL BE RELEASED AT LATER POINT IN COMPETITION*****)

Columns in .json file

- * **`category`** - "memes" type of content, for this dataset all are memes
- * **`img_captions`** - A literal description of what the image represents from an image captioning model like chatGPT.
- * **`meme_captions`** - (**GROUND TRUTH**)The description of what the meme tries to convey (**YOU HAVE TO PREDICT THIS**)
- * **`title`** - Title of meme that
- * **`url`** - Url of Meme
- * **`img_fname`** - "memes_<post_id>.png" use this to get images from "memes" folder
- * **`metaphors`** - Additional info
- * **`post_id`** - post_id

.